

引用格式:

段春燕, 王晓凌. 唐古特扁桃叶绿体基因组特征分析[J]. 湖南农业大学学报(自然科学版), 2024, 50(3): 27–34.

DUAN C Y, WANG X L. Characterization of chloroplast genome extracted from *Amygdalus tangutica*[J]. Journal of Hunan Agricultural University(Natural Sciences), 2024, 50(3): 27–34.

投稿网址: <http://xb.hunau.edu.cn>



## 唐古特扁桃叶绿体基因组特征分析

段春燕<sup>1</sup>, 王晓凌<sup>2</sup>

(1.河南科技大学农学院, 河南 洛阳 471023; 2.河南科技大学牡丹学院, 河南 洛阳 471023)

**摘要:** 采用 Illumina HiSeq X Ten 平台测序和生物信息法对唐古特扁桃叶绿体基因组序列特征进行解析。结果表明: 唐古特扁桃叶绿体基因组为四分体式结构, 完整叶绿体基因组序列全长 158 166 bp, G+C 含量 36.8%, 注释 131 个基因, 即 86 个蛋白编码基因、8 个 rRNA 基因和 37 个 tRNA 基因; 其叶绿体基因组共有 52 711 个编码密码子, 有 33 种类型是偏好密码子; 唐古特扁桃叶绿体基因组中 SSR 位点有 33 个, 串联重复有 13 个; SSR 中未检测到三核苷酸重复; SSR 分布不平衡, 大多位于基因间区 IGS 和 LSC 区域; 基于叶绿体全基因组序列构建系统发育建树, 发现唐古特扁桃与同亚属的榆叶梅、野樱桃以及李属毛樱桃亲缘关系较近。

**关键词:** 唐古特扁桃; 叶绿体基因组; 生物信息学; 密码子偏好性; 重复序列; 系统发育

中图分类号: R282

文献标志码: A

文章编号: 1007-1032(2024)03-0027-08

## Characterization of chloroplast genome extracted from *Amygdalus tangutica*

DUAN Chunyan<sup>1</sup>, WANG Xiaoling<sup>2</sup>

(1.College of Agriculture, Henan University of Science and Technology, Luoyang, Henan 471023, China; 2.Peony College, Henan University of Science and Technology, Luoyang, Henan 471023, China)

**Abstract:** Using the Illumina HiSeq X Ten sequencing platform and bioinformatics analysis methods, the sequence characteristics of the chloroplast genome of *Amygdalus tangutica* were analyzed in this study. The results showed that the whole chloroplast genome was a tetrad structure with 158 166 bp and G+C content of 36.8%. 131 genes were annotated, including 86 protein coding genes, 8 rRNA genes, and 37 tRNA genes. The protein coding genes encoded a total of 52 711 codons, of which 33 types were preferred codons. There were 33 SSRs and 13 tandem repeats in the chloroplast genome of *A. tangutica*. No trinucleotide repeats were detected in SSRs. The distribution of SSRs was not evenly, with most located in the IGS and LSC regions between genes. Based on the whole chloroplast genomic sequences, a phylogenetic tree was also constructed and analyzed. It was found that *A. tangutica* was closely related to *Amygdalus triloba* and *Amygdalus pedunculata* in the same subgenus *Amygdalus*, as well as *Prunus tomentosa*.

**Keywords:** *Amygdalus tangutica*; cpDNA(chloroplast genome); bioinformatics; codon preference; repeated sequence; phylogeny

唐古特扁桃(*Amygdalus tangutica*)又名西康扁桃、四川扁桃, 为蔷薇科(Rosaceae)扁桃(*Amygdalus* subgenus)亚属落叶灌木, 耐瘠薄、耐旱、耐寒, 为中国特有的优良木本油料植物<sup>[1]</sup>。唐古特扁桃油甾醇含量丰富, 超过 300 mg/(100 g), 具抗炎、抗癌、

降胆固醇和预防心脑血管疾病等生物活性<sup>[1-2]</sup>。其种子中苦杏仁甙含量丰富, 可用来治疗麻风病、哮喘、便秘和白斑病等多种疾病。其内果皮可加工制作活性炭; 种仁含油率高, 其油脂中富含甾醇、多酚、角鲨烯、维生素 E 和不饱和脂肪酸, 可开发高

收稿日期: 2023-08-24

修回日期: 2024-06-25

基金项目: 国家自然科学基金项目(U1304326); 河南科技大学博士科研启动基金资助项目(4024-13480114)

作者简介: 段春燕(1973—), 女, 河南沁阳人, 博士, 副教授, 主要从事植物分子系统学和植物资源学研究, 1445262292@qq.com

档食用油和生物柴油<sup>[1-3]</sup>。有关唐古特扁桃的研究主要集中在生物量估测模型<sup>[3]</sup>、资源评价<sup>[1]</sup>、生长繁殖<sup>[4]</sup>以及种仁油理化性质及脂肪酸组成<sup>[2]</sup>等方面,对唐古特扁桃的叶绿体基因组分子系统学研究尚少。本研究中,利用二代高通量技术完成唐古特扁桃的叶绿体基因组的测序,分析唐古特扁桃全叶绿体基因组特征,基于全叶绿体基因组数据进行系统发育分析,并与 NCBI 数据库(<https://www.ncbi.nlm.nih.gov/>)中其他近缘物种全叶绿体基因组进行比较,旨在为后续扁桃类植物的分子标记开发以及桃属植物等的分类和遗传多样性研究提供依据。

## 1 材料和方法

### 1.1 材料的采集和处理

采集唐古特扁桃植株的深绿色幼嫩叶片,分别以纯净水和干净面巾纸清洁表面,用变色硅胶颗粒干燥,超低温密封保存。

### 1.2 全基因组 DNA 的提取及测序

先后采用改良后的 CTAB 法<sup>[5-6]</sup>和德国 QIAGEN DNeasy Plant Mini Kit DNA 提取试剂盒提取样本总 DNA,经测序公司检测合格后,通过 Illumina HiSeq X Ten 双末端平台上机测序,测序深度为 10 倍。

### 1.3 叶绿体基因组组装及注释

测序原始数据以 FASTQ 格式存储,对低质量过滤后的数据通过 SOAPdenovo 2.04<sup>[7]</sup>在线拼接软件组装成重叠群,进一步优化组装结果<sup>[8-10]</sup>后采用 Gap Closer 对组装结果进行补洞和修正;结合 NCBI 上已报道的蔷薇科相关近缘物种的注释结果<sup>[11]</sup>,以 BLAST、DOGMA<sup>[12]</sup>和 Geneious R8<sup>[13-14]</sup> 软件进行 IR/SC 区的连接,基因注释,最后手动调整修正注释结果。使用 Sequin 软件提交测序相关数据到 NCBI。为表述方便,本研究中,叶绿体基因组和唐古特扁桃叶绿体基因组分别简称为 cpDNA 和 TC。利用 OGDRAW 在线工具绘制 TC 的环状图谱<sup>[15]</sup>。

### 1.4 叶绿体基因组遗传特征和系统发育分析

运用软件 Mega7.0<sup>[16]</sup>分析 TC 编码基因的密码子和 RSCU(相对同义密码子)使用情况,RSCU>1 表示该密码子有偏好性。

以在线软件 Tandem Repeats Finder 检索基因组中的串联重复(tandem repeats),分别设置比对参数匹配、错配和插入缺失为 2、7、7<sup>[8]</sup>;最小比值为 80 bp,最大比值为 500 bp。检测样本基因组序列中的 SSR(简单重复序列)位点,最小的 SSR 序列片段长度设置为 10 bp<sup>[8-10]</sup>,单、二、三、四、五和六核苷酸的最小重复次数各设置为 12、6、5、3、3 和 3,手动校正预测结果<sup>[9-10,17-18]</sup>。

IR/SC 区边界能够反映出被子植物不同物种间 cpDNA 的长度多态性。使用软件 IRscope 绘制 TC 与近缘物种 IR 边界对比图。以 Geneious 软件 Mauve 插件检测 TC 与近缘物种 cpDNA 的共线性情况<sup>[19-20]</sup>。将 TC 序列与从 GenBank 获得的其他 13 条 cpDNA 运用 MAFFT<sup>[21]</sup>进行序列比对,然后以 MEGA 7.0 通过邻接法(NJ)1 000 次抽样重复构建系统进化树。表 1 为本研究中的物种 cpDNA 序列信息。

表 1 本研究中的物种叶绿体基因组序列信息

Table 1 The chloroplast genome sequence information of the species in this study

物种	分类	登录号
唐古特扁桃( <i>Amygdalus tangutica</i> )	扁桃亚属	MK780039
榆叶梅( <i>Amygdalus triloba</i> )	扁桃亚属	MK790138
苹果( <i>Malus trilobata</i> )	苹果属	KX499858
野樱桃( <i>Amygdalus pedunculata</i> )	扁桃亚属	MG602257
樱桃( <i>Prunus pseudocerasus</i> )	李属	KX255667
蔷薇( <i>Rosa rugosa</i> )	蔷薇属	MN661138
山樱桃( <i>Amygdalus mongolica</i> )	扁桃亚属	KY073235
榆叶梅( <i>Amygdalus triloba</i> )	扁桃亚属	KY101155
矮扁桃( <i>Amygdalus tenella</i> )	扁桃亚属	NC-044965
大叶胡颓子( <i>Elaeagnus macrophylla</i> )	胡颓子属	KP211788
桃( <i>Amygdalus persica</i> )	扁桃亚属	HQ336405
扁桃( <i>Amygdalus dulcis</i> )	扁桃亚属	KY085904
扁桃( <i>Amygdalus dulcis</i> )	扁桃亚属	KY101152
野樱桃( <i>Amygdalus pedunculata</i> )	扁桃亚属	KY101153
毛樱桃( <i>Prunus tomentosa</i> )	李属	MF624726
山樱桃( <i>Amygdalus mongolica</i> )	扁桃亚属	MG602256
甘肃桃( <i>Amygdalus kansuensis</i> )	扁桃亚属	KF990036

## 2 结果与分析

### 2.1 叶绿体基因组测序和组装及注释情况

去掉低质量 reads 后,得到 Clean data,将高通量测序所得数据提交至 NCBI,获得登录号为 MK780039 的唐古特扁桃叶绿体基因组。TC 的序列总长度为 158 166 bp,为双链环型 DNA 分子,具有经典的四分体结构,由 2 个反向重复区域

(IRA/IRB)、1 个大单拷贝区域(LSC)和 1 个小单拷贝区域(SSC)组成(图 1)。其中, LSC、IRA、SSC

和 IRB 序列长度分别为 86 657、26 239、19 031、26 239 bp, TC 序列的 G+C 平均含量为 36.8%。

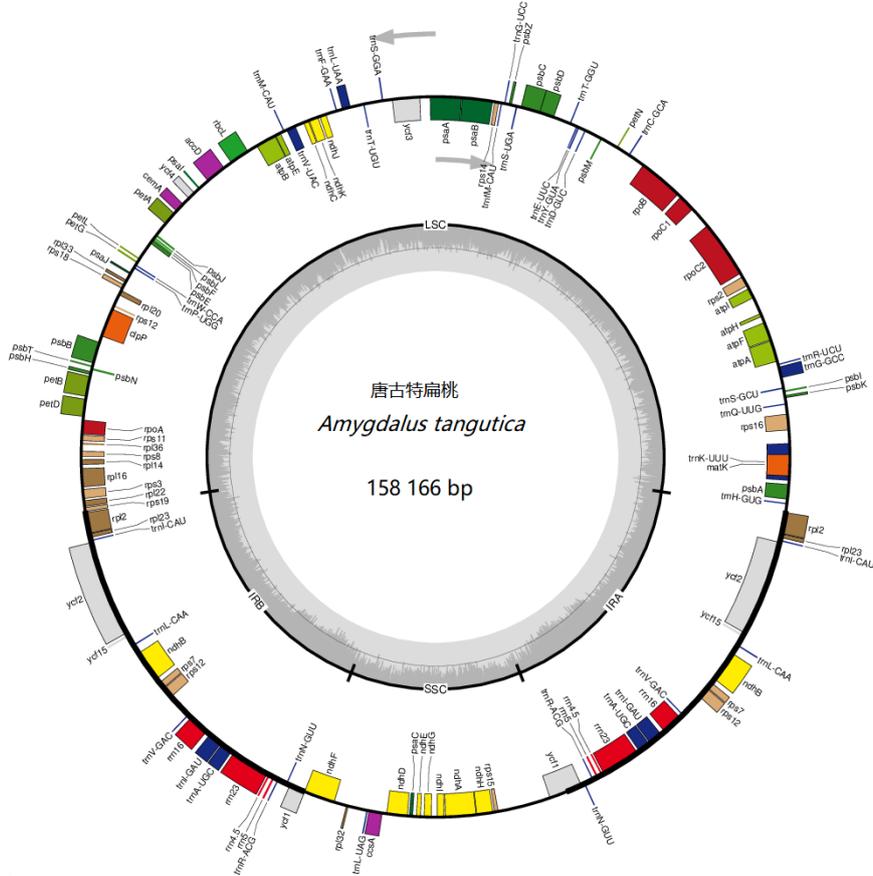


图1 唐古特扁桃叶绿体基因组四分体结构

Fig.1 Four segments in the chloroplast genome of *Amygdalus tangutica*

HQ336405、KX499858、MG602257 和 MN661138 分别对应桃 (*Amygdalus persica*)、苹果属 (*Malus trilobata*)、野樱桃 (*A. pedunculata*) 和蔷薇属 (*Rosa rugosa*) cpDNA 序列登录号(图 1)。TC 的长度与同属中的桃和野樱桃相近, 短于苹果的; 5 个物种中, LSC 长度均大于 SSC 与 IR。TC 的大单拷贝区长 86 657 bp, 小单拷贝区长 19 031 bp, IR 区长

26 239 bp(图 1)。TC 结构与蔷薇科的山樱桃、野樱桃、榆叶梅、蔷薇及其他被子植物相同, 具有典型的四分体结构: 由 LSC、IRA、IRB 和 SSC 不同区域构成。TC 总共编码 131 个基因, 包含编码蛋白基因 86 个, rRNA 基因 8 个(图 2, 表 2)。131 个基因分类和分组不同: 分类基因包括表达相关、光合作用相关、未知功能和其他等;



圈外侧表示正向编码的基因, 圈内侧表示反向编码的基因; 彩色条表示不同的功能组; 内部深灰色圈代表 G + C 占比; 浅灰色圈代表 A + T 占比。

图 2 唐古特扁桃叶绿体基因组图谱

Fig.2 Chloroplast genome map of *Amygdalus tangutica*

表2 唐古特扁桃叶绿体基因组注释情况

Table 2 Annotation of chloroplast genome of *Amygdalus tangutica*

分类基因	分组基因	基因名称
表达相关分类基因	核糖体 RNA 分组基因	<i>rrn4.5<sup>②</sup>, rrn5<sup>②</sup>, rrn16<sup>②</sup>, rrn23<sup>②</sup></i>
	转运 RNA 分组基因	<i>trnA-UGC<sup>**</sup>, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnI-M-CAU, trnG-GCC<sup>*</sup>, trnG-UCC, trnH-GUG, trnI-CAU<sup>②</sup>, trnI-GAU<sup>**②</sup>, trnK-UUU<sup>*</sup>, trnL-CAA<sup>②</sup>, trnL-UAA<sup>*</sup>, trnL-UAG, trnM-CAU, trnN-GUU<sup>②</sup>, trnP-UGG, trnQ-UUG, trnR-ACG<sup>②</sup>, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC<sup>②</sup>, trnV-UAC<sup>*</sup>, trnW-CCA, trnY-GUA</i>
光合作用相关分类基因	核糖体小亚基分组基因	<i>rps2, rps3, rps7<sup>②</sup>, rps8, rps11, rps12<sup>**②</sup>, rps14, rps15, rps16<sup>*</sup>, rps18, rps19</i>
	核糖体大亚基分组基因	<i>rpl2<sup>**②</sup>, rpl14, rpl16<sup>*</sup>, rpl20, rpl22, rpl23<sup>②</sup>, rpl32, rpl33, rpl36</i>
	RNA 聚合酶亚基分组基因	<i>rpoA, rpoB, rpoCI<sup>*</sup>, rpoC2</i>
	NADH 脱氢酶分组基因	<i>ndhA<sup>*</sup>, ndhB<sup>**②</sup>, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	光合系统 I 分组基因	<i>psaA, psaB, psaC, psaI, psaJ</i>
	光合系统 II 分组基因	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	细胞色素复合物分组基因	<i>petA, petB<sup>*</sup>, petD<sup>*</sup>, petG, petL, petN</i>
	ATP 合成酶分组基因	<i>atpA, atpB, atpE, atpF<sup>*</sup>, atpH, atpI</i>
	二磷酸核酮糖羧化酶大亚基分组基因	<i>rbcl</i>
	其他分类基因	成熟酶分组基因
	依赖 ATP 的蛋白酶单元 p 分组基因	<i>clpP<sup>**</sup></i>
	包裹膜蛋白分组基因	<i>cemA</i>
	乙酰辅酶 A 羧化酶亚基分组基因	<i>accD</i>
	c 型细胞色素合成分组基因	<i>ccsA</i>
未知功能分类基因	保守开放阅读框分组	<i>ycf1<sup>②</sup>, ycf2<sup>②</sup>, ycf3<sup>**</sup>, ycf4, ycf15<sup>②</sup></i>

\*代表该基因包含 1 个内含子; \*\*代表该基因包含 2 个内含子; ②代表该基因有 2 个重复单元。

分组基因种类包括核糖体 RNA、转运 RNA、核糖体大/小亚基、RNA 聚合酶亚基、NADH 脱氢酶、光合系统 I/II、细胞色素复合物、ATP 合成酶等多个类型。不同基因的序列长度有差异, TC 所包含基因中序列长度前 3 位的分别是: *ycf2* 长度为 6 834 bp, *rpoC2* 长度为 4 107 bp, *rpoB* 长度为 3 213 bp。TC 的 *rrn4.5*、*rrn5*、*rrn16*、*rrn23* 均为双拷贝, 位于 IR 区, 其基因长度分别为 103、121、1 491、2 809 bp, 基因长度与山樱桃(MG602256)和榆叶梅

(MK790138)的完全一致, 十分保守。

## 2.2 密码子偏好性分析

TC 序列共编码 52 711 个密码子。亮氨酸(L)、丝氨酸(S)和异亮氨酸(I)的编码密码子数量较多, 分别为 5 089、4 837 和 4 682 个, 编码亮氨酸的密码子最多, 占总量的 9.65%; 编码色氨酸(W)的密码子数量最少, 仅 697 个, 占总量的 1.32%(表 3)。唐古特扁桃叶绿体基因组中编码苯丙氨酸(F)的

表3 唐古特扁桃叶绿体基因组密码子编码氨基酸的数量和类型

Table 3 The numbers and types of amino acids encoded by the codon of the chloroplast genome of *Amygdalus tangutica*

密码子	数量	RSCU									
UUU(F)	2 348	1.24	UCU(S)	1 153	1.43	UAU(Y)	1 645	1.34	UGU(C)	782	1.26
UUC(F)	1 440	0.76	UCC(S)	934	1.16	UAC(Y)	805	0.66	UGC(C)	458	0.74
UUA(L)	1 171	1.38	UCA(S)	830	1.03	UAA(*)	1 332	1.31	UGA(*)	980	0.96
UUG(L)	1 047	1.23	UCG(S)	618	0.77	UAG(*)	745	0.73	UGG(W)	697	1.00
CUU(L)	1 029	1.21	CCU(P)	627	1.07	CAU(H)	1 001	1.43	CGU(R)	389	0.70
CUC(L)	672	0.79	CCC(P)	603	1.03	CAC(H)	403	0.57	CGC(R)	256	0.46
CUA(L)	676	0.80	CCA(P)	721	1.23	CAA(Q)	1 087	1.46	CGA(R)	567	1.03
CUG(L)	494	0.58	CCG(P)	395	0.67	CAG(Q)	399	0.54	CGG(R)	394	0.71
AUU(I)	1 926	1.23	ACU(T)	682	1.14	AAU(N)	1 879	1.37	AGU(S)	771	0.96
AUC(I)	1 162	0.74	ACC(T)	621	1.04	AAC(N)	855	0.63	AGC(S)	531	0.66
AUA(I)	1 594	1.02	ACA(T)	694	1.16	AAA(K)	2 131	1.38	AGA(R)	1 072	1.94
AUG(M)	844	1.00	ACG(T)	400	0.67	AAG(K)	954	0.62	AGG(R)	633	1.15
GUU(V)	781	1.34	GCU(A)	462	1.26	GAU(D)	998	1.42	GGU(G)	576	1.03
GUC(V)	430	0.74	GCC(A)	352	0.96	GAC(D)	404	0.58	GGC(G)	327	0.58
GUA(V)	714	1.23	GCA(A)	403	1.10	GAA(E)	1 232	1.35	GGA(G)	779	1.39
GUG(V)	402	0.69	GCG(A)	251	0.68	GAG(E)	594	0.65	GGG(G)	559	1.00

\*表示终止密码子。

UUU 数量最多,达 2 348 个,编码丙氨酸(A)的 GCG 数量最少,仅 251 个。此外,TC 中有 33 种密码子的相对同义密码子使用度(RSCU)>1,表明它们为偏好密码子。在偏好密码子中有 28 种密码子第 3 位是以 A 或者 U 结尾,占 RSCU 值大于 1 的密码子总量的 84.85%,具有明显的密码子第 3 位 A/U 选择偏好。

### 2.3 基因组重复序列以及 SSR 分析

根据设置参数,共识别出 13 个 TC 串联重复。

位于 LSC 的重复序列有 6 个,数量最多,占比为 46%;位于 IRA 和 IRB 的重复序列均为 3 个,占比均为 23%;位于 SSC 的重复序列只有 1 个,数量最少,占比为 8%。重复序列长度为 0~20 bp 的有 6 个,数量最多,占比为 46%;重复序列长度为 21~30 bp 的有 5 个,占比为 39%;数量最少的是重复序列长度>31 bp 的,只有 2 个,占比为 15%。植物基因组中的 SSR 具有高度多态性。在唐古特扁桃(*A. tangutica*)的 cpDNA 中,通过所设置的参数标准,共检测到 33 个 SSR 位点(表 4),包括 21 个

表 4 唐古特扁桃叶绿体基因组中的 SSR 情况

Table 4 Statistics of SSRs detected in the chloroplast genome of *Amygdalus tangutica*

序号	类型	SSR	大小/bp	起始点/bp	结束点/bp	所属基因	位置	区域间
1	p1	(T)17	17	1 693	1 709	<i>psbA/trnK-UUU</i>	LSC	IGS
2	c	(AAAT)3taattaaattattgatttaagtactttttct tattcttcccacaaccg(A)18	79	5 652	5 730	<i>rps16</i>	LSC	intron
3	p5	(TTTGA)3	15	6 069	6 083	<i>rps16</i>	LSC	intron
4	p1	(A)14	14	6 923	6 936	<i>rps16 / trnQ-UUG</i>	LSC	IGS
5	p1	(A)14	14	7 774	7 787	<i>psbK/psbI</i>	LSC	IGS
6	p4	(ATAA)4	16	8 674	8 689	<i>trnS-GCU / trnG-GCC</i>	LSC	IGS
7	p1	(T)13	13	10 004	10 016	<i>trnR-UCU / atpA</i>	LSC	IGS
8	p2	(AT)6	12	12 796	12 807	<i>atpF</i>	LSC	intron
9	p1	(T)12	12	13 178	13 189	<i>atpF</i>	LSC	intron
10	p4	(AAAG)3	12	13 337	13 348	<i>atpF/atpH</i>	LSC	intron
11	p1	(T)12	12	14 041	14 052	<i>atpH / atpI</i>	LSC	IGS
12	p1	(A)15	15	16 652	16 666	<i>rps2 / rpoC2</i>	LSC	IGS
13	p4	(AATA)3	12	29 642	29 653	<i>petN / psbM</i>	LSC	IGS
14	p4	(TTTA)3	12	31 014	31 025	<i>psbM/trnD-GUC</i>	LSC	IGS
15	p4	(TTTC)3	12	44 954	44 965	<i>ycf3</i>	LSC	intron
16	p1	(A)13	13	48 707	48 719	<i>trnT-UGU / trnL-UAA</i>	LSC	IGS
17	p2	(TA)7	14	50 545	50 558	<i>trnF-GAA/ndhJ</i>	LSC	IGS
18	p1	(T)14	14	58 926	58 939	<i>rbcL / accD</i>	LSC	IGS
19	p1	(C)12	12	60 838	60 849	<i>accD / psal</i>	LSC	IGS
20	p1	(A)12	12	65 364	65 375	<i>petA / psbJ</i>	LSC	IGS
21	p1	(G)13	13	66 553	66 565	<i>psbE / petL</i>	LSC	IGS
22	p1	(A)18	18	67 342	67 359	<i>psbE / petL</i>	LSC	IGS
23	p1	(T)15	15	69 516	69 530	<i>psaJ/rpl33</i>	LSC	IGS
24	p1	(A)13	13	70 250	70 262	<i>rps18</i>	LSC	CDS
25	p4	(TAAA)3	12	72 271	72 282	<i>clpP</i>	LSC	intron
26	p1	(T)17	17	72 417	72 433	<i>clpP</i>	LSC	intron
27	p1	(T)14	14	73 449	73 462	<i>clpP</i>	LSC	intron
28	p2	(AT)6	12	74 095	74 106	<i>clpP/psbB</i>	LSC	IGS
29	p1	(T)17	17	77 060	77 076	<i>psbH</i>	LSC	intron
30	p1	(T)19	19	84 679	84 697	<i>rpl16 / rps3</i>	LSC	IGS
31	p4	(TTGA)3	12	120 697	120 708	<i>ndhE</i>	SSC	CDS
32	p1	(T)12	12	122 175	122 186	<i>ndhI</i>	SSC	CDS
33	p1	(A)20	20	130 323	130 342	<i>ycfI</i>	SSC	CDS

p1 单核苷酸类型; p2 二核苷酸类型; p3 三核苷酸类型; p4 四核苷酸类型; p5 五核苷酸类型; c 复合类型; IGS、CDS和intron 分别表示该 SSR 位于叶绿体基因组的基因间区域、蛋白编码区域和内含子区域。

单核苷酸重复(63.6%), 3个二核苷酸重复(9.1%), 7个四核苷酸重复(21.2%), 1个五核苷酸重复(3.0%)和1个复合类型重复(3.0%), 并未检测到三核苷酸重复。在TC的21个单核苷酸重复中, 19个为AT类型, 占比达90.5%; 2个为GC类型, 占比较小, 仅9.5%。整个唐古特扁桃cpDNA中, SSR在各个区域中分布不均匀, LSC区域中有30个, SSC区域有3个, IR区域为0个; 在不同功能的基因间分布也不均匀, 在CDS蛋白编码区域有4个, 内含子区有10个, 基因间隔区有19个。从SSR分布的位置来看, SSR大多位于基因间区IGS, 占总数的57.58%。从碱基的组成上看, TC的SSR具有一定的A、T偏性, SSR类型组成主要是A或者T, 或AT复合重复。

## 2.4 IR/SC区边界的收缩与扩张分析

将唐古特扁桃与苹果、野樱桃、桃和蔷薇的叶绿体基因组序列进行IR/SC区边界的收缩与扩张分析(图3)。在植物叶绿体基因组中, LSC与IRB、IRB与SSC、SSC与IRA、IRA与LSC之间可依次形成JLB、JSB、JSA和JLA等4个边界。TC、MC、CC、

PC和RC分别代表唐古特扁桃叶绿体基因组(MK780039)、苹果叶绿体基因组(KX499858)、野樱桃叶绿体基因组(MG602257)、桃叶绿体基因组(HQ336405)和蔷薇叶绿体基因组(MN661138)。结果表明: 蔷薇科5种植物的叶绿体基因组IR(反向重复区域)长度在26 048~26 392 bp, 平均为26 289 bp; LSC(大单拷贝区域)为85 657~88 107 bp, 平均值为86 488 bp; SSC(小单拷贝区域)长度为18 780~19 316 bp, 平均值为19 043 bp; 总长度为156 533~160 207 bp, 平均值为158 109 bp。TC、MC、CC、PC和RC在IR/SC区域边界都有8个基因, 分别是*rpl22*、*rps19*和*rpl2*(LSC/IRB), *ycf1*和*ndhF*(IRB/SSC)、*ycf1*(SSC/IRA)和*rpl2*、*trnH*(IRA/LSC), 其中TC在JSB和JSA的*ycf1*和在JLA的*trnH*与其余4种植物的有明显差异。IRA区向*trnH*扩张的程度不同, MC、CC、PC和RC中*trnH*距离IRA/LSC边较近, 而TC的*trnH*距离IRA/LSC边较远。可见, 相对其他4种近缘植物, 唐古特扁桃的*trnH*向JLA边界方向的扩张幅度较小。

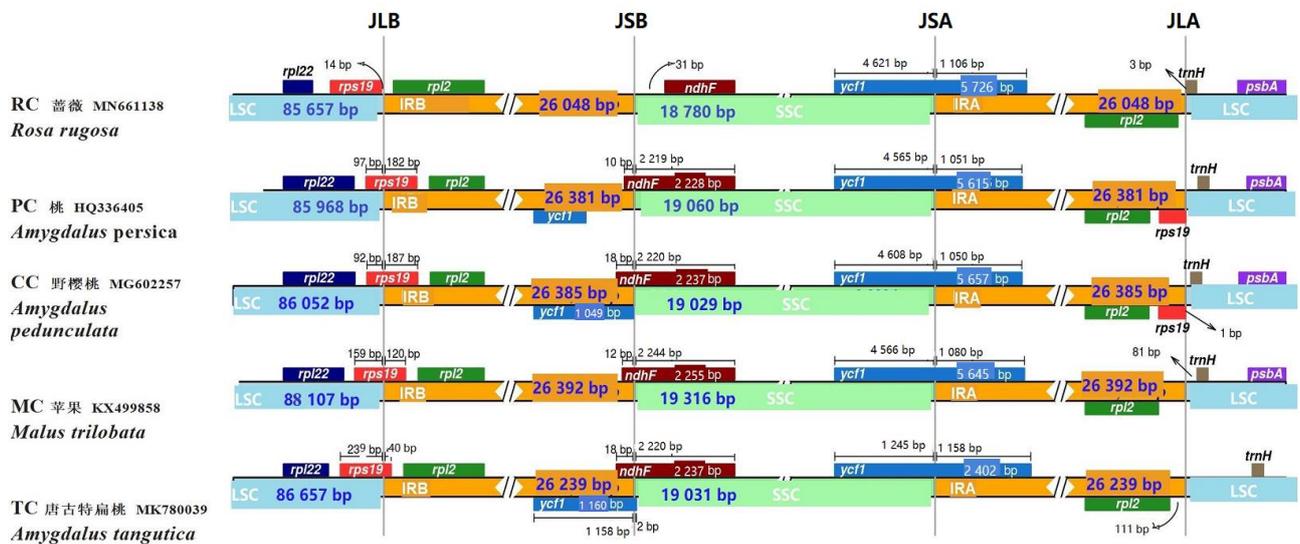


图3 唐古特扁桃的叶绿体基因组LSC、SSC和IR边界情况

Fig.3 The junction between LSC(large single copy-region), SSC(small single copy region) and IR(inverted repeat region) of the chloroplast genome of *Amygdalus tangutica*

## 2.5 基因组 Mauve 共线性分析

高等植物尤其是被子植物的cpDNA序列往往高度保守<sup>[18,22]</sup>, 由于DNA片段插入和删除, 基因组会出现重排现象。选择唐古特扁桃(MK780039)、苹果(KX499858)、野樱桃(MG602257)、桃(HQ336405)

和蔷薇(MN661138)等序列进行基因组全域Mauve比对。结果表明, 来自这5个物种的cpDNA序列均未检测到大片段的基因重排, 具备良好的cpDNA共线性关系。

## 2.6 系统发育分析

基于叶绿体基因组序列相关数据进行系统发育分析。选取唐古特扁桃(MK780039)、山樱桃(MG602256)、山樱桃(KY073235)、榆叶梅(KY101155)、矮扁桃(NC-044965)、扁桃(KY085904)、扁桃(KY101152)、野樱桃(KY101153)、毛樱桃(MF624726)、苹果(KX499858)、大叶胡颓子(KP211788)、樱桃(KX255667)、桃(HQ336405)、甘肃桃(KF990036),以 Mega7.0 软件构建进化树。结果(图 4)显示,在箭头区域唐古特扁桃和榆叶梅以高于 80% 的支持率聚于相同节点,毛樱桃与野樱桃、唐古特扁桃以 100% 的支持率分别聚于不同节点的同一大分支。可见,唐古特扁桃与毛樱桃、野樱桃和榆叶梅的亲缘关系最近,与桃和山樱桃亲缘关系稍远。在进行唐古特扁桃种质资源的野外保护和开发、品种培育以及生产实践时可考虑以榆叶梅、毛樱桃和野樱桃作为唐古特扁桃的砧木。

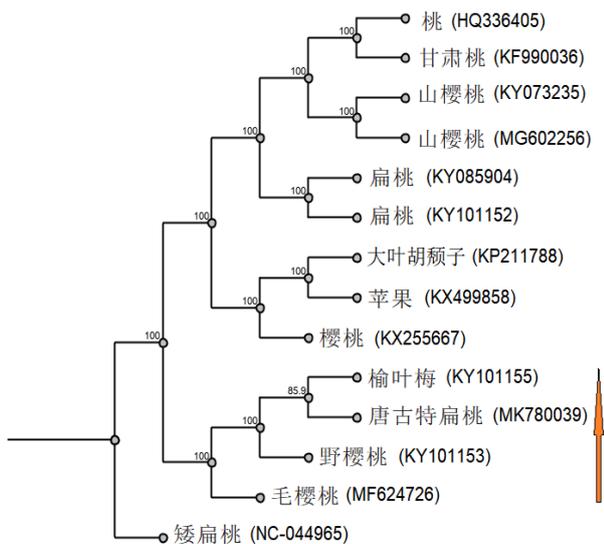


图 4 唐古特扁桃及其近缘种 NJ 系统发育树

Fig.4 The phylogenetic tree constructed based on the chloroplast genomes of *Amygdalus tangutica* and some related species using NJ method

## 3 结论与讨论

叶绿体为质体的一种,具半自主性,是绿色植物进行光合作用的场所<sup>[22-24]</sup>。cpDNA 多为母性遗传,比核基因组相对分子质量小且突变率较低,具有序列组成和基因排列保守的特点。被子植物的 cpDNA 一般包括拷贝区 SC 和重复区 IR,具四分体

结构。被子植物的 cpDNA 编码基因数多在 130 个左右,序列总长度常小于  $1.60 \times 10^5$  bp<sup>[22-26]</sup>。本研究中,对唐古特扁桃叶绿体基因组进行测序、注释和分析,发现唐古特扁桃的 cpDNA 序列总长度为 158 166 bp,具有经典的四分体结构,为双链环型 DNA 分子,符合一般高等植物 cpDNA 的结构和大小范围。有研究<sup>[27-28]</sup>表明,扁桃亚属中矮扁桃为二倍体原始种,野樱桃为十二倍体,属于年轻物种。本研究发现,唐古特扁桃 cpDNA 序列总长度介于矮扁桃(158 596 bp)<sup>[11]</sup>和野樱桃(157 851 bp)之间。3 种扁桃植物的 cpDNA 序列平均长度 158 204 bp,最小的是野樱桃(MG602257),最大的是矮扁桃(MK764428),唐古特扁桃 cpDNA 序列总长度较矮扁桃和野樱桃序列总长度的差值分别为 430 bp 和 315 bp,更接近于野樱桃。

生物体内遗传信息传递需要密码子对应承担,密码子是 DNA 转录与翻译、蛋白质合成与表达过程中的中介。密码子常具有简并性,即 2~6 个密码子对应编码 1 种氨基酸。同义密码子存在一定使用次数和频率的偏好性。密码子偏好性是生物体长期适应环境和进化中的结果<sup>[13,19,22]</sup>。唐古特扁桃 cpDNA 中偏好密码子有 33 种,且偏好密码子的第 3 个核苷酸具有明显的 A/U 偏好。

本研究通过 Illumina HiSeq X Ten 双末端测序平台获得唐古特扁桃叶绿体基因组序列,以叶绿体基因组数据进行系统发育分析,唐古特扁桃和榆叶梅以高于 80% 的支持率聚于相同节点,野樱桃与唐古特扁桃以 100% 的支持率聚于同一大分支,表明唐古特扁桃与同亚属中的野樱桃和榆叶梅亲缘关系相近,此结论与通过植物性状和核 DNA 分析研究结果<sup>[11,19-20,27-28]</sup>基本一致。

## 参考文献:

- [1] 王伟,许新桥.木本油料西康扁桃资源评价及产业化发展策略[J].中国油脂,2016,41(11):1-5.
- [2] 柴春山,莫保儒,蔡国军,等.甘肃迭部野生西康扁桃种仁油理化性质及脂肪酸组成分析[J].中国粮油学报,2023,38(6):98-105.
- [3] 韦丹丹,包维楷.岷江干旱河谷西康扁桃地上生物量估测模型[J].应用与环境生物学报,2021,27(2):315-324.
- [4] 韦丹丹,刘鑫,王子龙,等.岷江上游干旱河谷西康扁桃灌丛生长繁殖特征[J].植物科学学报,2021,39(4):389-397.

- [5] YANG J B, LI D Z, LI H T. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs[J]. *Molecular Ecology Resources*, 2014, 14(5): 1024–1031.
- [6] 宋艳波, 吴国良, 牛洪斌. 改良 CTAB 法在核桃叶片基因组 DNA 提取中的应用研究[J]. *山西农业大学学报(自然科学版)*, 2011, 31(2): 109–112.
- [7] LUO R B, LIU B H, XIE Y L, et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler[J]. *Giga Science*, 2012, 1: 18.
- [8] 段春燕, 王晓凌. 蒙古扁桃叶绿体遗传特征分析[J]. *北方园艺*, 2024(2): 31–37.
- [9] 童一涵, 郑倩, 杜新明, 等. 多齿红山茶叶绿体基因组序列特征分析[J]. *植物资源与环境学报*, 2022, 31(5): 27–36.
- [10] 段春燕. 中国桃属-扁桃亚属植物全叶绿体基因组比较分析和系统发育研究[D]. 西安: 西北大学, 2020.
- [11] 杨斌, 孟庆瑶, 张凯, 等. 孑遗濒危植物矮扁桃叶绿体全基因组特征分析及亲缘关系鉴定[J]. *植物研究*, 2020, 40(5): 686–695.
- [12] WYMAN S K, JANSEN R K, BOORE J L. Automatic annotation of organellar genomes with DOGMA[J]. *Bioinformatics*, 2004, 20(17): 3252–3255.
- [13] DUAN C Y, SHEN Y H, ZHAO G F. Complete chloroplast genome characteristics of *Prunus triloba* Lindl[J]. *Mitochondrial DNA Part B, Resources*, 2020, 5(1): 504–505.
- [14] KEARSE M, MOIR R, WILSON A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data[J]. *Bioinformatics*, 2012, 28(12): 1647–1649.
- [15] GREINER S, LEHWARK P, BOCK R. Organellar GenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes[J]. *Nucleic Acids Research*, 2019, 47(W1): W59–W64.
- [16] KUMAR S, STECHER G, TAMURA K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. *Mol Biol Evol*, 2016, 33: 1870.
- [17] 侯哲, 娄晓鸣, 李昂, 等. 星毛唐松草叶绿体基因组特征及系统发育分析[J]. *植物资源与环境学报*, 2023, 32(3): 24–32.
- [18] BEIER S, THIEL T, MÜNCH T, et al. MISA-web: a web server for microsatellite prediction[J]. *Bioinformatics*, 2017, 33(16): 2583–2585.
- [19] 段春燕, 张凯, 段义忠. 长柄扁桃叶绿体基因组比较[J]. *生物工程学报*, 2020, 36(12): 2850–2859.
- [20] 段春燕, 王晓凌. 重瓣榆叶梅全叶绿体基因组遗传特征分析[J]. *浙江农林大学学报*, 2024, 41(3): 577–585.
- [21] KATO H K, STANDLEY D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. *Molecular Biology and Evolution*, 2013, 30(4): 772–780.
- [22] 钱方, 高作敏, 胡利娟, 等. 海甘蓝(*Crambe abyssinica*) 叶绿体基因组特征及其系统发育研究[J]. *生物技术通报*, 2022, 38(6): 174–186.
- [23] 丁玉梅, 杨正安, 周晓罡, 等. *Bt* 基因转入烟草叶绿体获得转化植株的研究[J]. *湖南农业大学学报(自然科学版)*, 2007, 33(S1): 189–192.
- [24] 陈咸吉, 聂现辉, 艾文胜, 等. 基于叶绿体 DNA 条形码的竹子种属聚类分析与耐盐性鉴定[J]. *湖南农业大学学报(自然科学版)*, 2021, 47(5): 514–522.
- [25] 龚意辉, 周桂花, 李丽梅, 等. 黄桃叶绿体基因组的组装与序列分析[J]. *热带作物学报*, 2022, 43(8): 1545–1553.
- [26] 富贵, 刘晶, 李军乔. 密花香薷叶绿体基因组结构及系统进化分析[J]. *中草药*, 2022, 53(6): 1844–1853.
- [27] 邱蓉, 程中平, 王章利. 中国扁桃亚属植物亲缘关系及其演化途径研究[J]. *园艺学报*, 2012, 39(2): 205–214.
- [28] 董山平. 基于染色体核型和单拷贝核基因 DNA 序列的榆叶梅及其近缘种系统学研究[D]. 北京: 北京林业大学, 2015.

责任编辑: 毛友纯  
英文编辑: 柳正